IBM - Applied Data Science Capstone

Capstone Project - The Battle of Neighbourhoods

DATA COLLECTION

• Latitude and longitude coordinates of those neighbourhoods. This is required in order to plot the map.

• Venue data, particularly data related to shopping malls. We will use this data to perform clustering on the neighbourhoods.

• The geographical coordinates of the neighbourhoods are collected using Python Geocoder package which will give us the latitude and longitude coordinates of the neighbourhoods.

• The details for the mall location is scrapped from the FourSquare developers website.

the data was sourced from various websites, csv files and government datas in repositories. and then cleaned and combined for optimal use.

data used.

1. the data provided by the indiapost organisation (in csv form)
2. the data provided by the government of india (https://data.gov.in/catalog/all-india-pincode-directory?filters%5Bfield_catalog_reference%5D=85840&format=json&offset=0&limit=6&sort%5Bcreated%5D=desc)
3. the geolocation data provide by indiapost and kaggle.com
4. webscrapped data from ("https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Coimbatore")

hence all the data is combined with the commonality in single feature called as pincode/postalcode.

hence india is diversified nation and also higly populated every region has lots of subcategory and neighborhood. hence it is mandatory to get deep down and link all the regions.

the following attributes are collected from this data sources,

2.1 Neighborhood(3 levels of sub categories in region)

2.2 Latitude

2.3 Longitude

2.4 Venue Name

2.5 Venue Latitude

2.6 Venue Longitude

2.6 Venue Category